

Remote Sensing Information Gateway (RSIG)

Overview of Key Features & Architecture

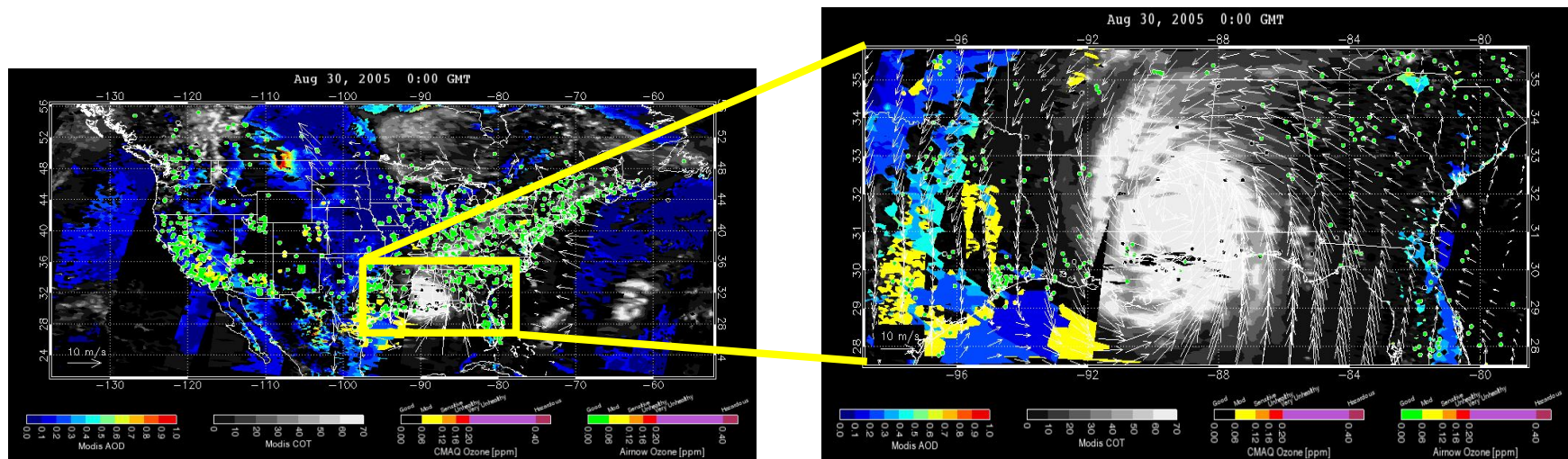
RSIG Key Features

- **Subsetting huge files at the source and saving the reduced dataset to local file(s), respecting accepted standards and conventions**
 - Saving time and enhancing convenience
- **Single website virtually serving many kinds of data**
 - Only need an EPA Portal account to access the applet, which reads and displays data from multiple sites
- **Aggregation**
 - Combine data from thousands of files to a single data stream
- **Visualization**
 - Interactive data selection and visualization
- **Fast**
 - Compare 50+ hours of downloading satellite files (before RSIG) to approximately 5 minutes (using RSIG)

RSIG Key Features

➤ Subsetting

- Time: extract & transfer only data within hourly range (over many days)
- Variable: extract & transfer only variables of interest
- Domain: extract & transfer only data within a chosen longitude-latitude box



For example, each (daily met) CMAQ file is 442 rows x 265 columns x 22 layers x 10 variables = 5GB

The subsetting file is only 2.5MB—1/2000th the source data size. Convenient, fast, and efficient to transmit across the network.

RSIG Key Features

- **Save subsetted data to a local file in various formats**
 - **XDR binary: Simple human-readable ASCII header followed by simple portable binary data arrays (easily read by model applications)**

```
AIRNOW 1.0
Full US Domain
2004-07-22T00:00:00-0000
# data dimensions: timesteps stations
120 482
# Variable names:
PM25
# Variable units:
ug/m3
# IEEE-754 32-bit reals sites[stations][2=<longitude,latitude>] and
# IEEE-754 32-bit reals data[timesteps][stations]:
```

- **Tab-delimited ASCII: Simple to import into a spreadsheet program**

Timestamp (UTC)	LONGITUDE(deg)	LATITUDE(deg)	PM25 (ug/m3)
2004-07-22T00:00:00-0000	-91.0561	30.7008	20.0
2004-07-22T00:00:00-0000	-90.2728	30.0411	-999.0
2004-07-22T00:00:00-0000	-90.8114	30.3150	12.0

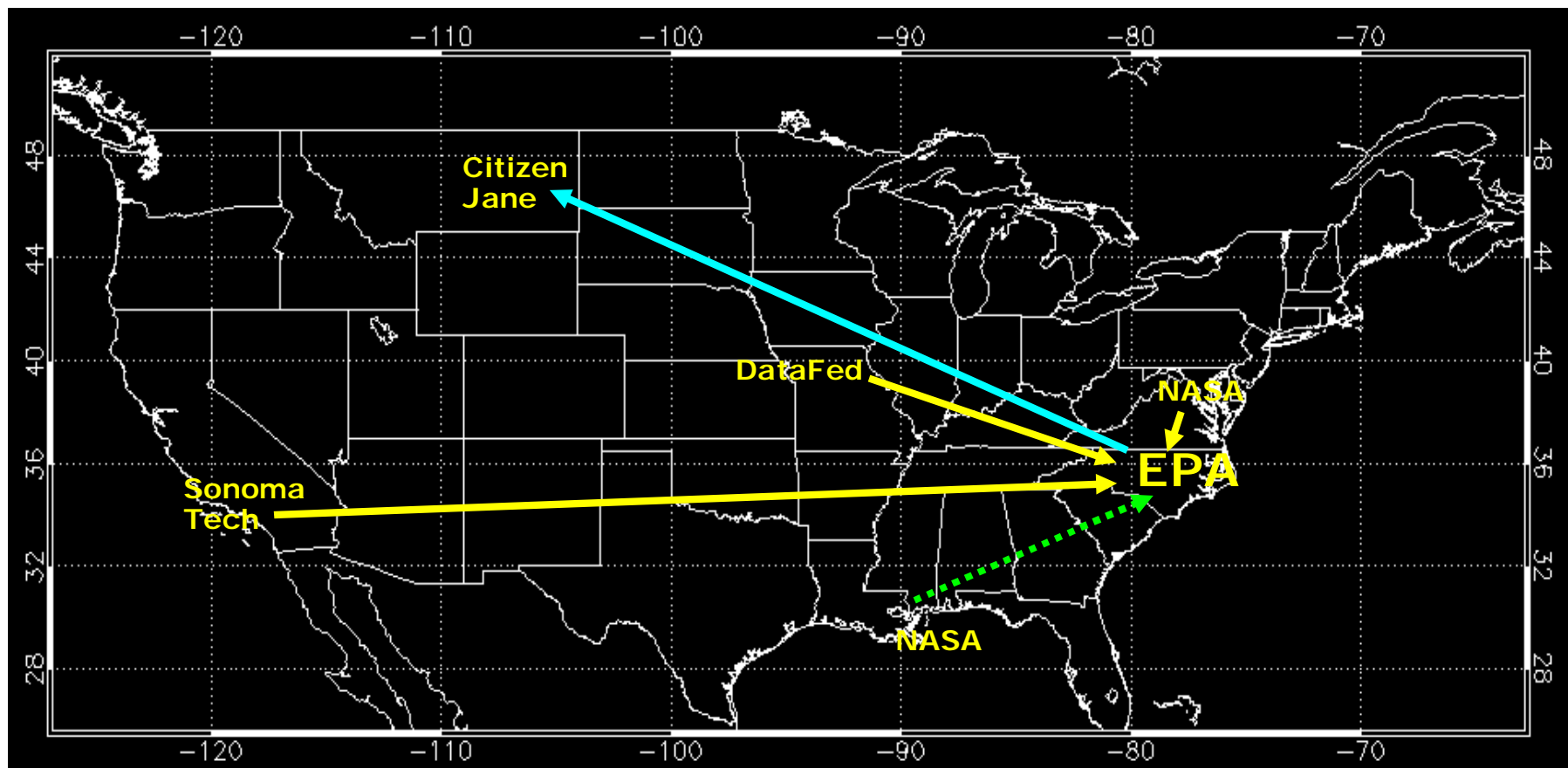
RSIG Key Features

- **Save subsetting data to a local file in various formats**
 - **HDF:** Original satellite data files for those who need them
 - **PNG, MPEG, KMZ** formats for images/movies/Google Earth
 - **GeoTIFF, ARC Grid** and other requested formats can be added
 - **NetCDF:** CMAQ/M3IO conventions or COARDS, CDF, and so on

```
netcdf modis {  
  dimensions:  
    points = 169222 ;  
  variables:  
    float longitude(points) ;  
      longitude:units = "degrees_east" ;  
    float latitude(points) ;  
      latitude:units = "degrees_north" ;  
    double Scan_Start_Time(points) ;  
      Scan_Start_Time:units = "seconds since 1993-1-1 00:00:00.0 0" ;  
    float Cloud_Optical_Thickness(points) ;  
      Cloud_Optical_Thickness:units = "none" ;  
      Cloud_Optical_Thickness:missing_value = -9999.f ;  
    int yyyyddd(points) ;  
      yyyyddd:units = "date" ;  
    int hhmmss(points) ;  
      hhmmss:units = "time" ;  
    float time(points) ;  
      time:units = "hours since 2005-08-26 03:25:00.0 -00:00" ;  
  // global attributes:  
    :west_bound = -90.f ;  
    :east_bound = -85.f ;  
    :south_bound = 30.f ;  
    :north_bound = 35.f ;  
    :Conventions = "COARDS" ;  
    :history = "http://daac.gsfc.nasa.gov/MODIS/Terra/  
product_descriptions_modis.shtml#atmos -> MODISSubset -> XDR2COARDS" ;  
}
```

RSIG Key Features

- **Single website virtually serving many kinds of data**
 - CMAQ model gridded air-quality and meteorology (Ozone, PM2.5, Temperature, Wind, etc.)
 - AIRNow/AQS ground station measurements (ozone, PM)
 - MODIS satellite sensed data (Ozone, Aerosol Optical Depth, Clouds, etc.)
 - CALIPSO satellite LIDAR data (backscatter, etc.)
 - NESDIS Biomass Burning surface values (CO, PM2.5, etc.)



RSIG Key Features

➤ Aggregation

- Combine data *from thousands of files* covering the subset of interest *into a single data stream*

For example, 10 days of unsubsetted data occupies 200GB of disk space. MODIS data alone counts for about 3,500 files:

```
sapphire:/project/rsig/demo/data:52 >du -sk.  
186730490.
```

```
sapphire:/project/rsig/demo/data:53 >ls -l MODIS/* | wc -l 3465
```

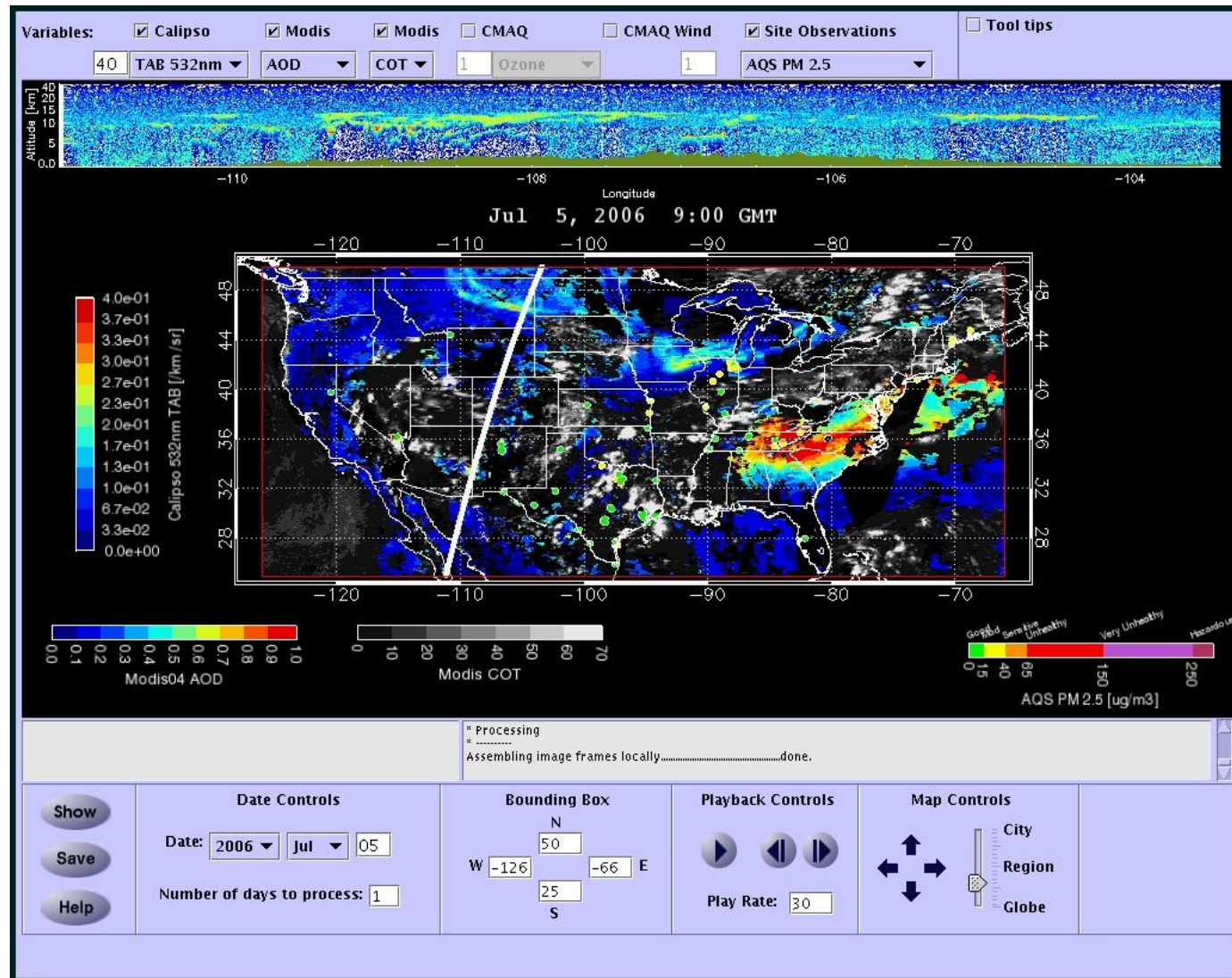
RSIG quickly determines which few files might cover the region of interest:

```
sapphire:/project/rsig/demo/data/MODIS/2005:59 >head bounds.mod04 ; echo "... " ; tail bounds.mod04  
MOD04_L2.A2005237.0000.004.2005237130801.hdf    147.668506   175.317569   15.309453    36.463558  
MOD04_L2.A2005237.0005.004.2005237125733.hdf    144.281166   169.226692   -2.459210    18.520328  
MOD04_L2.A2005237.0010.004.2005237130934.hdf    139.782254   164.963577   -20.355334    0.630548  
MOD04_L2.A2005237.0015.004.2005237130807.hdf    133.320187   161.766645   -38.259256   -17.076903  
MOD04_L2.A2005237.0020.004.2005237125713.hdf    122.526061   159.414321   -56.189316   -34.394028  
MOD04_L2.A2005237.0025.004.2005237130252.hdf    100.268832   160.528160   -74.027835   -50.942405  
MOD04_L2.A2005237.0120.004.2005237130416.hdf    -180.000000   180.000000    66.539196    89.997721  
MOD04_L2.A2005237.0125.004.2005237130431.hdf     - 80.000000   180.000000    61.180824    86.201825  
MOD04_L2.A2005237.0130.004.2005237130901.hdf    126.647362   174.841691    45.827809    68.277003  
MOD04_L2.A2005237.0135.004.2005237130344.hdf    125.000007   157.798333    28.891350    50.370785  
...
```


RSIG Key Features

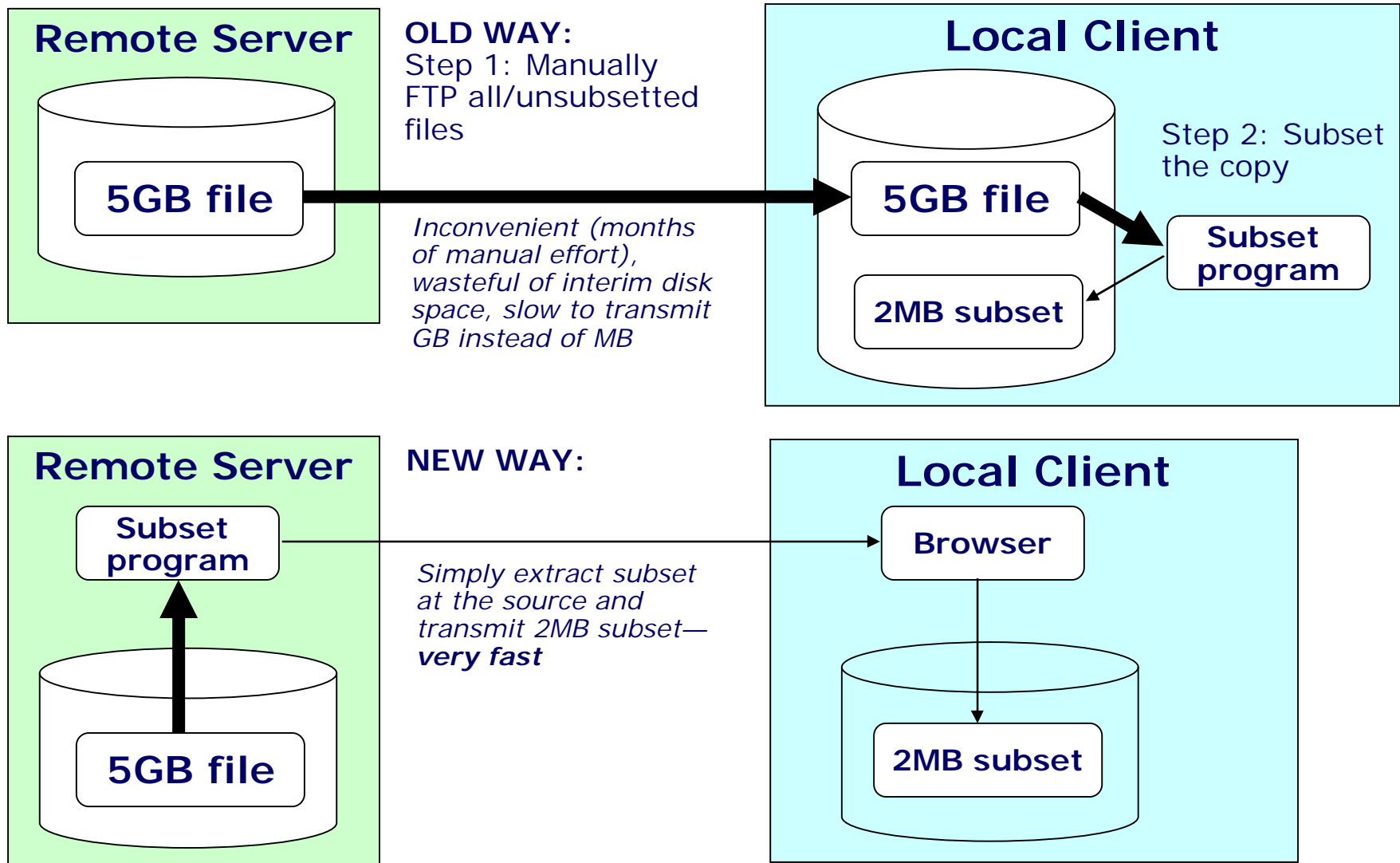
➤ Visualization

- Interactive animation (30 frames/second) of all data on a map displayed in a web browser. User can save animation as MPEG file.



RSIG Key Features

- **Fast:** *Months* of manual effort to retrieve data can be done in *seconds*



Getting Data – the OLD way

- **To get 6 days of CALIPSO data for the continental USA, here's what to do.**
 - Register with the DAAC and create an account.
 - Log in to the Web site.
 - Select “project” and specify the parameters—*after* you've understood the NASA nomenclature
 - **Wait three days** for the email alerting you that the files you want have been demigrated from tape
 - FTP the HDF files. For this request, that means 144 files of about 500MB each, totaling about 726GB. This will take **over 50 hours** to download.
 - Process the HDF files into a usable format. Only 23 files of those 144 intersect the continental USA – it's up to you to sort them out.

Back
Forward

[Login](#)
[New User](#)
[Update Info](#)
[Retrieve Login & Password](#)
[CD-ROM & Videocassette](#)
[Tool Unavailability](#)
Please check out the *Tool Unavailability* page to see scheduled down times for this tool.

SEARCH FILES

Projects

CALIPSO

Total: 1

Parameters

AEROSOL OPTICAL THICKNESS
AEROSOL PARTICLE PROPERTIES
BACKSCATTER COEFFICIENT
BAROMETRIC PRESSURE

Total: 29

Refine Lists

Reset Lists

Files Data Sets [Get Info](#)

16028 CAL_LIR_L1-PROV-V1-10
4782 CAL_LID_L1-PROV-V1-10
1804 CAL_LID_L1-PROV-V1-11
4553 CAL_LID_L1-PROV-V1-20
2440 CAL_LID_L1-PROV-V1-22
15939 CAL_LID_L1-PROV-V2-01
6253 CAL_LID_L2_01kmCLay-PROV-V1-10
6857 CAL_LID_L2_01kmCLay-PROV-V1-20
8221 CAL_LID_L2_01kmCLay-PROV-V2-01

Total: 26

ADVANCED SEARCH

Time Ranges	Start Date	Stop Date
Optional (YYYY-MM-DD)	1993-01-01	2008-02-05
Geographic Search	Top 90.00	
	Left -180.00	180.00 Right
	-90.00 Bottom	

Day/Night: ☒ Both ☐ Day ☐ Night

Clear Form

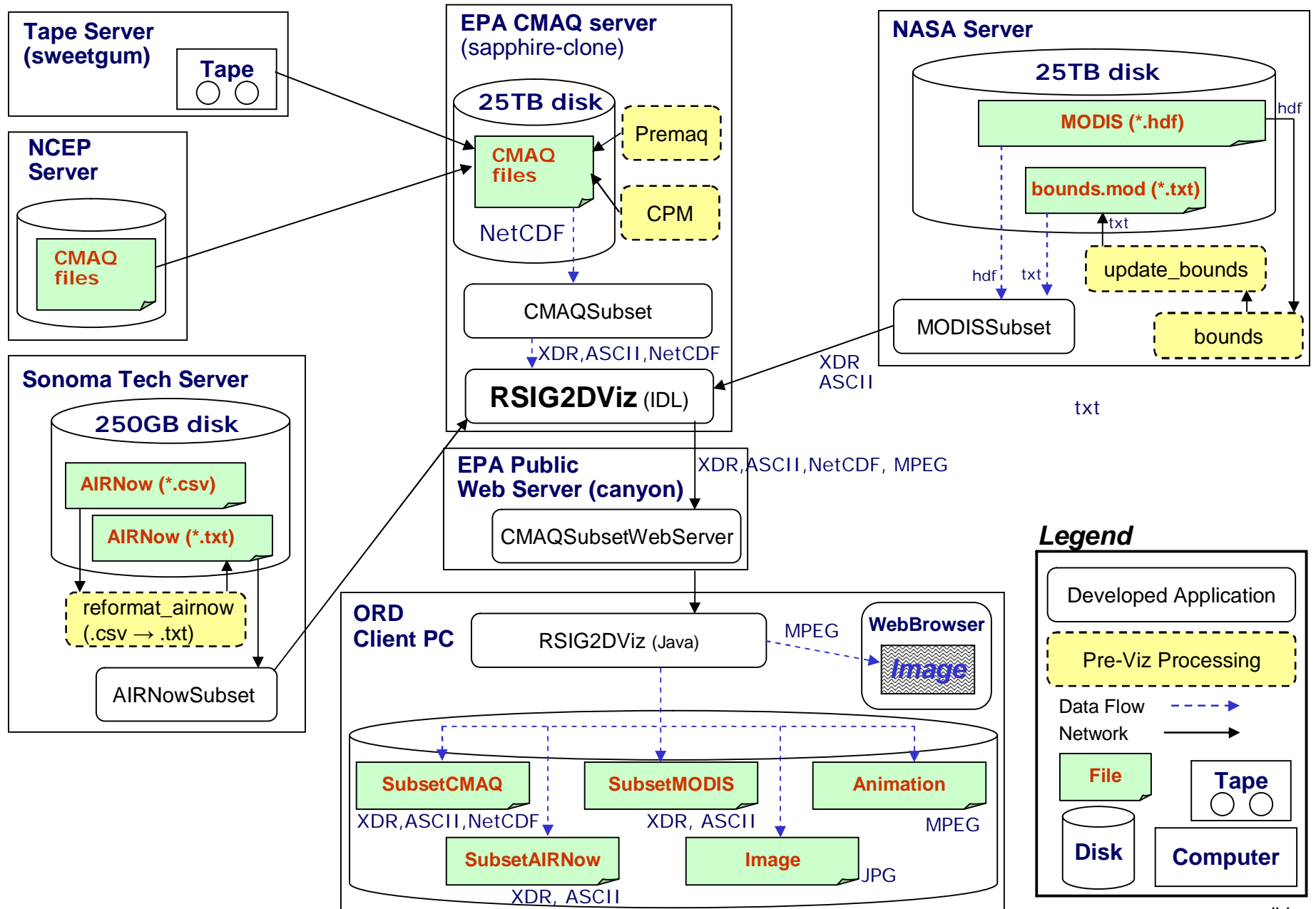
Get Results

slide 11

Getting Data – the RSIG way

- **To get 6 days of CALIPSO data for the continental USA, here's what to do.**
 - **Start the browser applet.**
 - **Select the CALIPSO dataset.**
 - **Draw a bounding box over the continental USA, enter the starting date, and enter “6” to specify the 6 days you want.**
 - **Wait about 5 minutes** for the low-res visualization, or, if you are streaming the full-resolution data, wait **about 5 minutes**.
 - **Save the data in any of several popular, standard formats.**
 - **Go back to work.**

RSIG System Data Flow Diagram



Data Visualization & Retrieval—The RSIG Way

- **Visualization occurs at a rate of “about a minute per day.”**
 - **It took 5 minutes to stream and visualize 6 days worth of data, with the following datasets:**
 - **CALIPSO LIDAR Backscatter**
 - **MODIS Aerosol Optical Depth**
 - **MODIS Cloud Optical Thickness**
 - **AIRNow PM 2.5**
 - **NESDIS Biomass Burning PM 2.5**

Data Visualization & Retrieval—The RSIG Way

- **Saving the data subset to local disk**
 - 5 minutes to stream and save full-resolution subsetting data
 - Yielded under 250MB in compressed simple-format files.

```
$ date
Thu Feb 7 15:52:48 EST 2008
$ date
Thu Feb 7 15:57:56 EST 2008
$ ls -asl
  64 -rw-rw-r-- 1 plessel visstaff      31978 Feb 7 15:55 airnow_PM2.5_20071023-6.xdr.gz
    0 -rw-rw-r-- 1 plessel visstaff      0 Feb 7 15:55 aqs_PM2.5_20071023-6.xdr.gz
473272 -rw-rw-r-- 1 plessel visstaff 242314947 Feb 7 15:55 calipso_TAB532nm_20071023-6.xdr.gz
   16 -rw-rw-r-- 1 plessel visstaff    5319 Feb 7 15:55 goes-bb_PM2.5_20071023-6.xdr.gz
69000 -rw-rw-r-- 1 plessel visstaff 35298901 Feb 7 15:58 modis2_COT_20071023-6.xdr.gz
 9904 -rw-rw-r-- 1 plessel visstaff 5070301 Feb 7 15:55 modis_AOD_20071023-6.xdr.gz
```

- **How long would this have taken using the "web form + email + ftp" method?**
- **How many files would need to be downloaded?**
- **How many GB of local disk space would be needed to store the files?**

Data Visualization & Retrieval—The RSIG Way

➤ Uncompressed XDR files total 1GB.

```
$ unzip *.gz
$ ls -asl
 224 -rw-rw-r-- 1 plessel visstaff    114367 Feb  7 15:55 airnow_PM2.5_20071023-6.xdr
1615560 -rw-rw-r-- 1 plessel visstaff 827165389 Feb  7 15:55 calipso_TAB532nm_20071023-6.xdr
  128 -rw-rw-r-- 1 plessel visstaff    61444 Feb  7 15:55 goes-bb_PM2.5_20071023-6.xdr
292400 -rw-rw-r-- 1 plessel visstaff 149704919 Feb  7 15:58 modis2_COT_20071023-6.xdr
 36232 -rw-rw-r-- 1 plessel visstaff   18548851 Feb  7 15:55 modis_AOD_20071023-6.xdr
```

➤ ASCII headers and XDR binary arrays are easy to read/parse.

```
$ head -12 airnow_PM2.5_20071023-6.xdr
AIRNOW 1.0
subset
2007-10-23T00:00:00-0000
# data dimensions: timesteps stations
144 194
# Variable names:
pm25
# Variable units:
ug/m3
# MSB 32-bit integers ids[stations] and
# IEEE-754 32-bit reals sites[stations][2=<longitude,latitude>] and
# IEEE-754 32-bit reals data[timesteps][stations]:
```

Example IDL code to Read AIRNow XDR stream

```
command = 'wget -q -c -T 0 -O - ' + "'" + $
          'http://maple.rtpnc.epa.gov/cgi-bin/airnowserver?' + $
          'SERVICE=wcs&VERSION=1.0.0&REQUEST=GetCoverage&' + $
          'COVERAGE=pm25&' + $
          'TIME=2005-08-26T20:00:00Z/2005-08-27T01:59:59Z&' + $
          'BBOX=-90,30,-88,35,0,0&FORMAT=xdr' + "'"

spawn, command, unit = pipe

; Read ASCII header of XDR-format data stream:

header = strarr( 12 )
line = ''

for index = 0, 11 do begin
    readf, pipe, line
    header[ index ] = line
Endfor

timestamp = header[ 2 ] ; E.g., 2007-10-23T00:00:00-0000.
variable = header[ 6 ] ; E.g., pm25.
units = header[ 8 ] ; E.g., ug/m3.

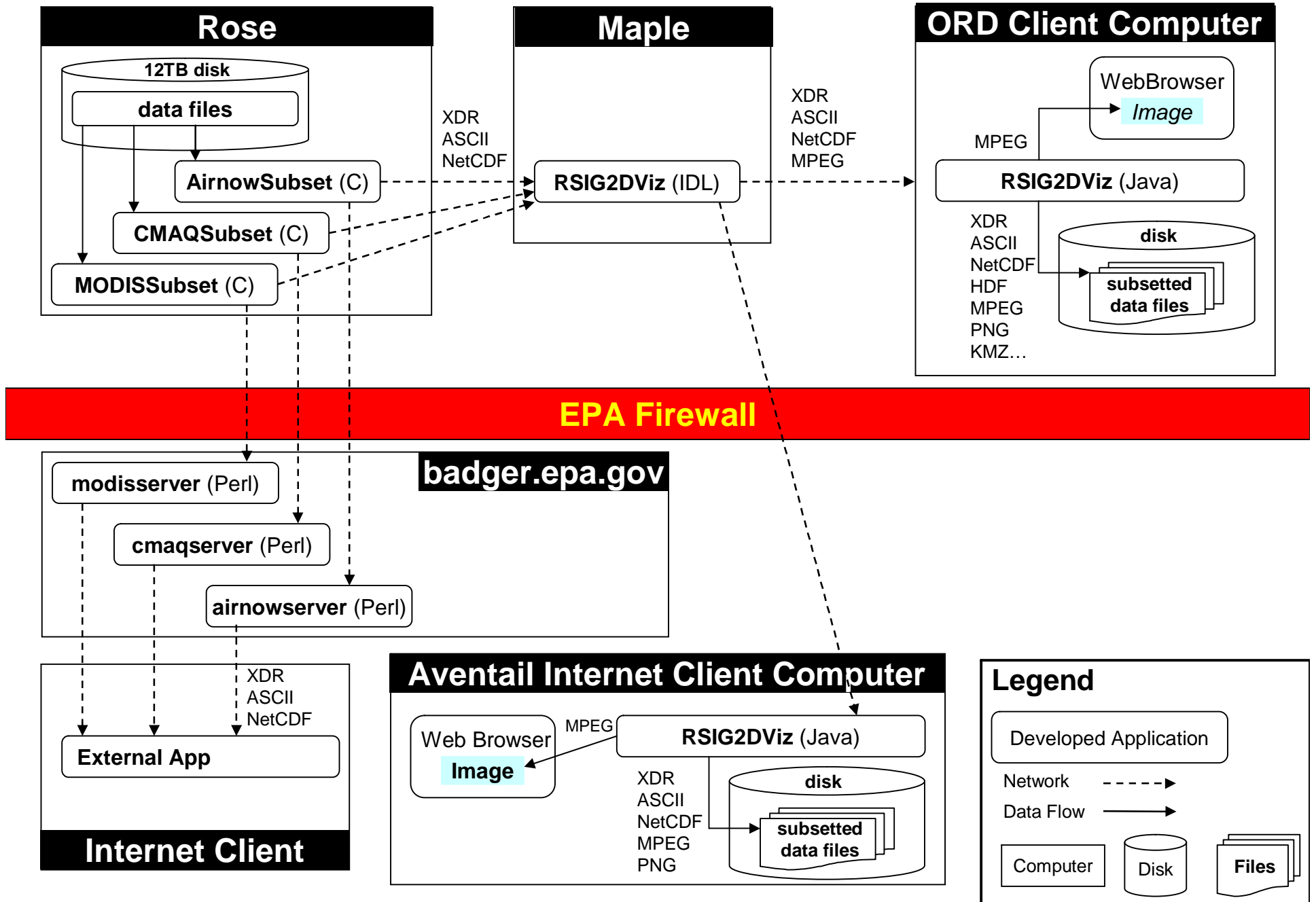
; Allocate data arrays:

dimensions = strsplit( header[ 4 ], ' ', /extract )
timesteps = long( dimensions[ 0 ] ) ; E.g., 144.
stations = long( dimensions[ 1 ] ) ; E.g., 194.
ids = lonarr( stations )
lonlats = fltarr( 2, stations ) ; IDL multi-dimensional arrays are column-major like FORTRAN...
data = fltarr( stations, timesteps )

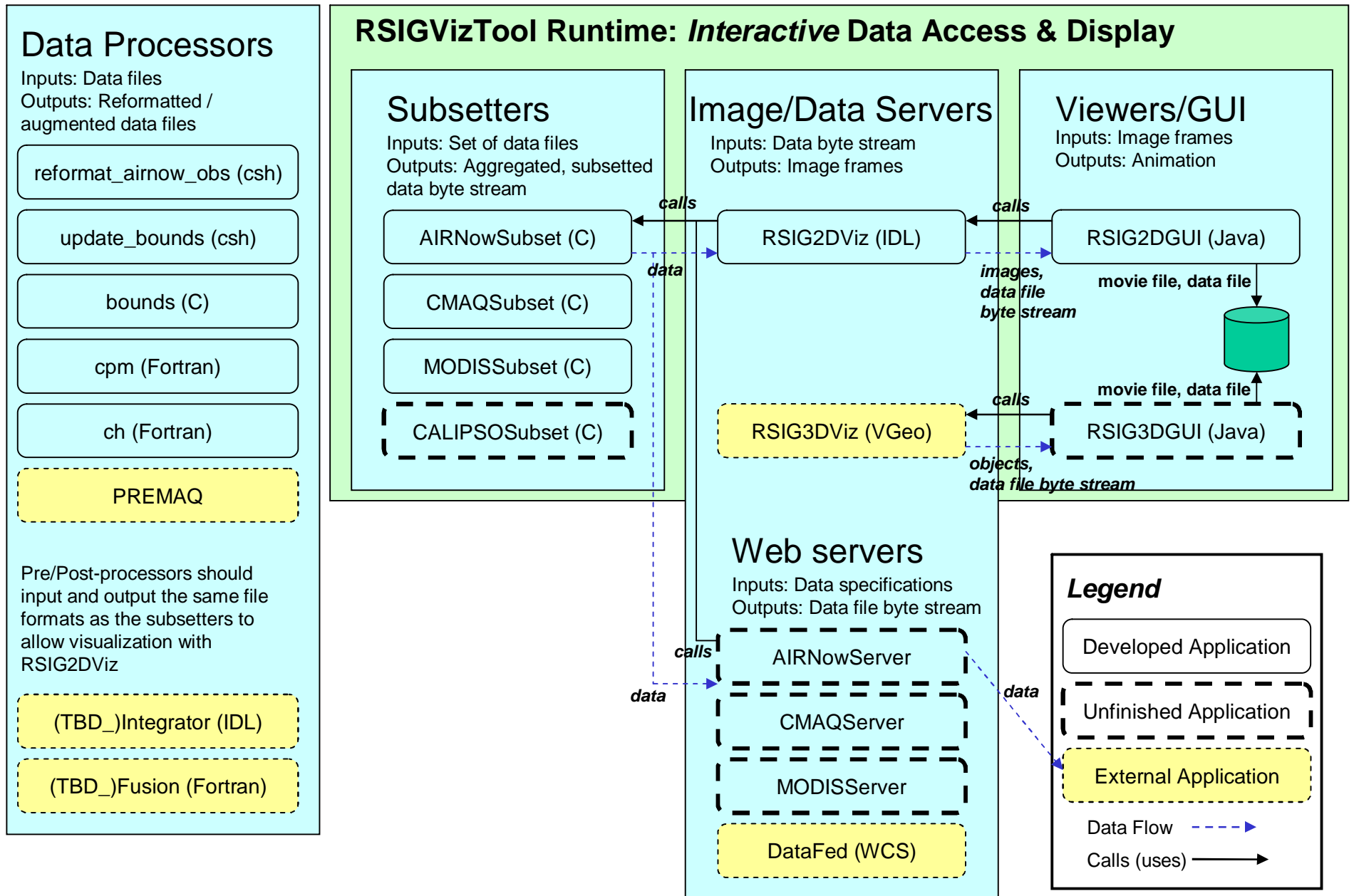
; Read XDR binary data arrays:

readu, pipe, ids
readu, pipe, lonlats
readu, pipe, data
swap_endian_inplace, ids, /swap_if_little_endian
swap_endian_inplace, lonlats, /swap_if_little_endian
swap_endian_inplace, data, /swap_if_little_endian
close, pipe
free_lun, pipe
result = { timestamp: timestamp, variable: variable, units: units, $
          timesteps: timesteps, stations: stations, ids: ids, lonlats: lonlats, data: data }
print, result
```

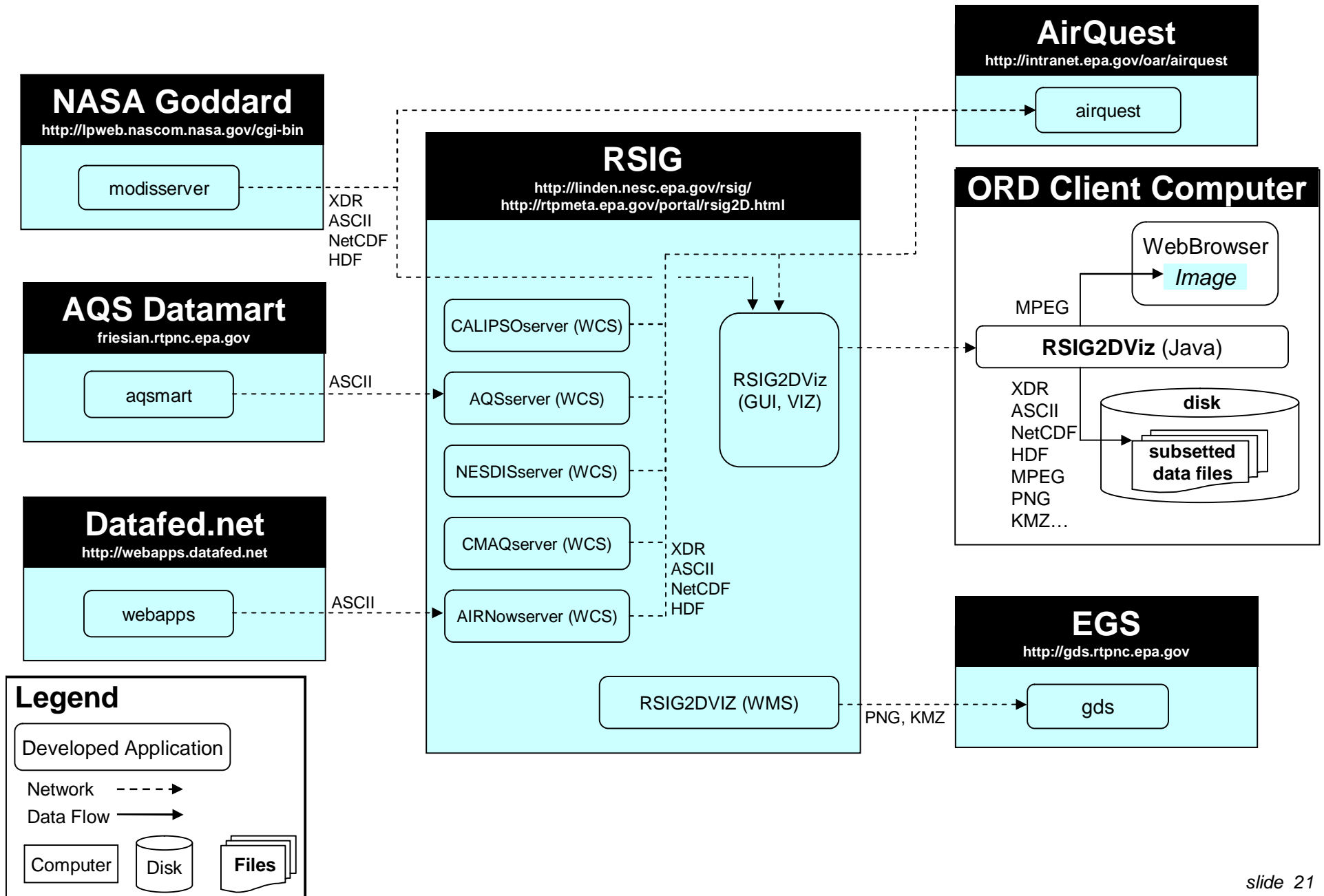
Current RSIG Runtime Data Flow



RSIG System Component Diagram



Connections To and From RSIG



Interoperability Data Standards Supported by RSIG – 1 of 2

- **OGC-WCS/XML**
 - Obtain description of available data (variables, time-range,...)
- **OGC-WCS**
 - Specify data subset by variable, time, lon-lat box.
- **OGC-WMS**
 - Specify and receive images and movies.
- **ASCII**
 - Human-readable format for small amounts of data streamed from WCS.
- **IEEE-754**
 - Computer-readable format for real-valued data streamed from WCS.
- **XDR**
 - Efficient streamable computer-readable format for real and integer data.
- **NetCDF**
 - Widely-used file format for arrays of scientific data.
- **NetCDF/COARDS**
 - NetCDF conventions for geo-spatial data.
- **NetCDF/IOAPI**
 - NetCDF conventions for gridded model met and AQ data.

Interoperability Data Standards Supported by RSIG – 2 of 2

- **HDF**
 - Widely used file format for arrays of scientific data.
- **HDF-EOS**
 - HDF conventions used by NASA.
- **PNG**
 - Most portable image file format.
- **MPEG**
 - Most portable movie file format.
- **KML/KMZ**
 - Geo-referenced image/movie format used by Google Earth and similar applications.
- **HTML**
 - Used for web-based documentation.
- **PDF**
 - Portable Document Format for fancy formatted content.
- **HTTP/80**
 - Communication protocol for web-based applications.
- **SI**
 - Mother of all standards. RSIG converts data to SI units. (F to C, mb to Pa, western longitudes are negative, UTC, etc.)

Why Standards are Not Enough

- In practice, interoperability requires more than just compliance with existing standards. The devil is in the details.
- Following are some gotchas we had to overcome.

Why Standards are Not Enough

- OGC–WCS DescribeCoverage and GetCoverage omits data units (!)
- Workaround: RSIG includes data units in its WCS label so users can get at the data themselves.

`<label>ozone(ppb)</label>`

`<label>pm25(ug/m3)</label>`

`<label>pres(Pa)</label>`

`<label>co2(metric tons)</label>`

Why Standards are Not Enough

- **WCS does not nail down the specifics of the format of the resulting data, for example, NetCDF-COARDS vs NetCDF-IOAPI, CSV, etc.**
 - **Universal data models *do not exist*.**
- **HDF/4/5/EOS: NASA satellite data files use inconsistent conventions for describing/encoding/structuring data, such as for the bounding box, variable names, dimensionality, data order, non-SI units or even no units!, SDS, VData**

Why Standards are Not Enough

➤ Performance issues:

- Too slow if data files are on tape or if the database has no appropriate index file
- Need formats that are fast, efficient, simple and streamable (i.e., no disk I/O), such as RSIG's custom ASCII/XDR format vs NetCDF
- Custom Extensions to WCS: STRIDE=2, SPARSE=800, COMPRESS=1

➤ Units (SI), time (UTC), missing data value sentinels, etc.

Planned Enhancements

- Install WCS servers, web pages and downloadable applet on badger.epa.gov (EPA Portal) to allow access from computers outside the EPA network.
- WCS servers will be modified with an option to regrid (project and aggregate) all data onto CMAQ grid cells. This will be done at the source by custom-developed XDRConvert program for efficiency and performance.
- Add new data sources including: MOPITT CO, AURA-OMI NO₂, SCIAMACHY NO₂, GASP, etc.
- Add new output formats: GeoTiff, ESRI/ARC Grid, IOAPI (regrid), etc.
- Launch HB/MCMC and read and display resulting data from HB and LARCTM models, etc.

Areas of Focus

- Access from non-EPA computers for research partners using the EPA Portal.
- Availability of EPA data outside EPA, such as CMAQ “base case” runs.
- Availability, performance, and convenient access to external data via WCS.
- Use RSIG WCS/WMS servers in other applications such as AirQuest and Environmental GeoWeb Server (EGS). EMVL can assist other developers with this.

END
February 8, 2008